Impacts on efficiency of merging the Swedish district courts

Per J. Agrell[1] · Pontus Mattsson[2] · Jonas Månsson[2]

Abstract

Judicial courts form a stringent example of public services using partially sticky inputs and outputs with heterogeneous quality. Notwithstanding, governments internationally are striving to improve the efficiency of and diminish the budget spent on court systems. Frontier methods such as data envelopment analysis are sometimes used in investigations of structural changes in the form of mergers. This essay reviews the methods used to evaluate the *ex post* efficiency of horizontal mergers. Identification of impacts is difficult. Therefore, three analytical frameworks are applied: (1) a technical efficiency comparison over time, (2) a metafrontier approach among mergers and non-mergers, and (3) a conditional difference-in-differences approach where non-merged twins of the actual mergers are identified by matching. In addition, both time heterogeneity and sources of efficiency change are examined *ex post*. The method is applied to evaluate the impact on efficiency of merging the Swedish district courts from 95 to 48 between 2000 and 2009. Whereas the stated ambition for the mergers was to improve efficiency, no structured *ex post* analysis has been done. Swedish courts are shown to improve efficiency from merging. In addition to the particular application, this work may inform a more general discussion on public service efficiency measurement under structural changes, and their limits and potential.

✉ Pontus Mattsson
  pontus.mattsson@lnu.se

  Per J. Agrell
  per.agrell@uclouvain.be

  Jonas Månsson
  jonas.mansson@lnu.se

[1] Center of Operations Research and Econometrics, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

[2] Department of Economics and Statistics, Linnaeus University, 351 95 Växjö, Sweden

1 Introduction

More effective utilization of public services is a general target for governments. Whereas the demand for increased capacity and quality of performed services may be infinite, the scarcity of allocated resources through public funds or direct fees forces a critical review of the efficiency of the service provision. Focusing here on specific services, such as judicial courts and governmental agencies (regulators), their assessment is particularly difficult for two reasons. First, the measurement of output may be challenging both in terms of quantification (aggregation) and in terms of quality dimensions. Second, the inputs, both in terms of senior staff (e.g. judges) and assets (e.g. court houses) may be of fixed or at least semi-fixed (sticky) character as, for example, argued by Ouellette and Vierstraete (2010). This means that adjustments to actually performed output may be slow or non-existent. Thus, the conventional means of efficiency improvement (managerial incentives, lawn-moving, budget reallocations) may not be applicable or effective. However, an instrument that is frequently used is reorganizing the service through horizontal mergers, i.e., closing courts or agencies and transferring their competencies to other existing units. The *ex ante* arguments presented in this regard include economies of scale, economies of scope, improved central coordination, improved operational risk sharing and ultimately, higher efficiency and lower cost (Bogetoft and Wang 2005).[1] Furthermore, OECD (2011) states that *ex post* evaluations of mergers is an important tool for reviewing previous decisions and creating future improvements.

Previous literature on evaluations of merging effects on performance either adopt a cost approach to evaluate whether cost savings occurred (e.g., Schmitt 2017) or investigate technical efficiency using data envelopment analysis (DEA) (e.g., Ferrier and Valdmanis 2004). Efficiency effects of merging may take time (Kwoka and Pollitt 2010) or the advantageous effects may never occur due to, for example, cultural dissimilarities (Cartwright and Cooper 1993). Consensus as to whether merging is beneficial can, therefore, not be found (e.g., Avkiran 1999; Garden and Ralston 1999; Ferrier and Valdmanis 2004).

Turning to court efficiency, the topic has been both of European[2] and national[3] interest. The main intervention for efficiency enhancements over the last decades has been to merge. However, to the best of our knowledge no structured *ex post* study has been conducted.

This paper applies a non-parametric evaluation model to assess the *ex post* effects on technical efficiency as a result of recent horizontal mergers among the Swedish district courts. Specifically, we address four questions: (1) what are the effects on efficiency of mergers, (2) are the effects of mergers temporal or stationary in time, (3), what are the driving factors of observed efficiency changes in courts and (4) investigation of the individual mergers in comparison to *ex ante* potential gains obtained by the Bogetoft and Wang (2005) model. The particular research questions relate directly to national policy (Eliasson et al. 2017; Ministry of Finance 2017), aiming at dimensioning the court system for potential output expansion. Moreover, we claim that the method and the empirics may also provide valuable information on horizontal mergers in public services beyond the court system.

Identification of reliable effects of a policy intervention is, in general, problematic, especially in the case of small samples. Therefore, our approach evaluates the question using three different analytical frameworks. First, the technical efficiency scores calculated by DEA are

---

[1] For a description of internal arguments to merge in the scope of Swedish district courts, see Mattsson and Tidanå (2019).

[2] See the creation, the European Commission leading to creation of 'The European Commission for the Efficiency of Justice' (CEPEJ).

[3] The Swedish Budget Act (SFS 2011: 203) explicitly states that public services need to be provided with a high level of efficiency.

compared over time using a global frontier, including the sources of change. Second, the merged and non-merged groups are compared in relation to organizational and managerial efficiency using a metafrontier approach. The global frontier and the metafrontier approaches are non-standard policy evaluation tools and used as support for the results in our third and major approach to estimate impacts of a policy intervention, i.e. a conditional difference-in-differences (DiD) approach. For this part of the analysis, merged courts are matched to non-merged courts using the Mahalanobis distance metric. Finally, we investigate the factor-level sources based on changes in inputs and outputs. The data for the application is obtained from the Swedish National Court Administration (SNCA) and consists of all Swedish district courts from 2000 to 2017, including the identity and scope for all structural changes.

The structure of the paper is as follows: Sect. 2 provides an overview of the Swedish judicial system targeting the district courts and its reorganization. Section 3 reviews previous literature, both on general approaches for assessing the efficiency of courts and the methodology for analyzing merging effects. Furthermore, Sect. 4 describes the application separated into three parts, i.e. technical efficiency, identification, and data. Section 5 presents the results. Finally, Sect. 6 concludes and provides policy recommendations.

2 The Swedish judicial system and mergers of the district courts in Sweden

The Ministry of Justice is responsible for the administration of the judicial system (Ministry of Justice 2015). The aim of the Swedish legal justice system is to provide fair trials, meaning that the Ministry is not allowed to interfere in the actual exercise of law. This guarantees independence and autonomy of courts, in relation to the Parliament, Government, and other executive branches. In a democratic society, courts are of fundamental importance, meaning that they have a special position compared to other institutions and authorities (Ministry of Justice 2000). There are three instances of general courts in Sweden, where the lowest one are the district courts, the second Courts of Appeal, and, the highest is the Supreme Court.

Swedish district courts
District courts were initiated in their present form in 1971. They mainly handle cases related to their local jurisdiction area, corresponding to the surrounding geographical area. However, there are also five courts specializing in land and environmental cases. These courts deal with e.g. environmental and water issues, property registration, and building matters. Within each court, there are Chief Judges, Senior Judges, and Judges who are considered as permanent judges. Furthermore, law clerks are judges in training who mainly work with case preparation. Finally, other staff works in support functions, such as human resource management (Ministry of Justice 2015).

SNCA separates court cases into three main categories: (1) criminal cases, (2) civil cases and (3) petitionary matters. Criminal cases are brought to justice by the prosecutor on behalf of the state or an individual. Civil cases are legal disputes between two or more parties, referred to a court for settlement in accordance to civil law or a contract. Last, petitionary matters are subject to a summary process regulated in the Court Matters Act (SFS 1996:242) and separated into four categories by SNCA: (1) debt clearances; (2) debt enforcements; (3) bankruptcies or company reconstructions; and (4) other matters.[4]

---

[4] Other matters are, for example, for example: custodians, parking tickets, heritages, and fiduciaries.

Table 1 Summary of mergers of Swedish district courts (Detailed information is presented in Table 7 in the "Appendix")

Year	Mergers	Number of merged courts
2000	2	4
2001	9	24
2002	3	9
2004	4	8
2005	7	18
2006	2	4
2007	5	7
2009	4	9
All	36	83

Reorganization of the district courts

In 1971 there were 108 district courts, decreasing to 95 by the year 2000 (Wångmar 2003). The reorganizations after 2000 have mainly occurred through mergers. Internal arguments for the mergers are, for example, that it would be easier to recruit qualified personnel to courts in larger cities and that a strengthening of the drafting organization would enable specialization (Ministry of Justice 2000). This has resulted in 36 mergers including two or more district courts during the period 2000–2009, summarized in Table 1.

From Table 1 it can be observed that most mergers took place in 2001 and in 2005 when no fewer than 42 courts merged. Over the period, mergers took place all over Sweden with the exception of 2007, when district courts in the Stockholm area were restructured. These courts did not follow a clear merging process in the sense that two or more courts were merged. Instead, the new courts consist of parts of the initial court.

During the period of 2000–2009, the number of district courts decreased from 95 to 48, which has stayed constant until 2018. In 2017, courts that merged during this period were approximately twice as large as non-merged courts.[5] Typically, a merger consisted of a relatively large district court taking over one or more smaller adjacent district court.

3 Previous research

Efficiency and total factor productivity (TFP) within courts have previously been investigated in several studies starting with Lewin et al. (1982) and Kittelsen and Førsund (1992), respectively. Less attention has been given to mergers. To the best of our knowledge, only Finocchiaro Castro and Guccio (2016) investigate the potential, *ex ante*, gains of merging within the scope of district courts and Santos and Amado (2014) examine whether small courts have a higher degree of inefficiency than larger due to planned absorption. Additionally, Falavigna et al. (2018) investigate whether efficiency could be enhanced by reducing the number of sections in the courts. However, no study has been found that analyzes the *ex post* efficiency results of mergers among district courts. Therefore, in the next section of the literature review, we turn our attention to model specifications for efficiency studies of courts, and in the last section we discuss general approaches used to assess merger effects on efficiency and cost.

---

[5] This is measured as the total number of employees, described in more detail in Sect. 4.3.

3.1 Inputs and outputs in measuring efficiency for courts

The majority of investigations regarding efficiency in district courts use the non-parametric technique DEA offered by Charnes et al. (1978). Lewin et al. (1982), and all other studies found, include the number of employees as an input. Employees can be measured as the number of judges (Ferrandino 2014; Finocchiaro Castro and Guccio 2014; Falavigna et al. 2018), administrative staff (Silva 2018) or as a separation between judges and administrative personnel (Pedraja-Chaparro and Salinas-Jimenez 1996; Santos and Amado 2014; Major 2015).[6] Caseload, consisting of pending and new cases, is included as an input in Schneider (2005) and Nissi and Rapposelli (2010), based on the argument that performance will be underestimated if the workload is a binding upper bound on utilization efficiency. This is, however, only an effect of the sticky nature of the inputs, because a lower workload can be compensated by less input to achieve higher efficiency, as argued by Mattsson et al. (2018). Using caseload as an input may also introduce bias based on poor performance (through backlogging of demand) and/or lack of resources in previous periods (Santos and Amado 2014). Inputs related to fixed capital are omitted in most studies, with the exception of Elbialy and García-Rubio (2011), who used the number of computers as a proxy for capital.

Outputs are generally measured as the number of closed cases (Nissi and Rapposelli 2010; Falavigna et al. 2018), which are, in some studies, separated by class; e.g., criminal and civil cases (Finocchiaro Castro and Guccio 2016). Further analysis of the relative complexity of the cases within classes is normally absent, resorting to simple counting. An exception is Santos and Amado (2014), who use duration-based weight restrictions among 43 classes of cases.

Finally, heterogeneous quality among courts may affect the assessment. Court delay has been used as a proxy for service quality in Falavigna et al. (2015), modeled as an undesirable output. However, service quality is more frequently analyzed through second-stage regression or correlation approaches. Examples of investigated quality variables are judges' salaries and educational level (Schneider 2005; Deyneli 2012).

3.2 Merging effects on costs and efficiency

Only a limited number of methodological approaches have been presented to analyze the *ex post* efficiency of mergers. A rare parametric approach is found in Çelen (2013) applying the Battese and Coelli (1992) stochastic frontier model to capture the merging effect by including a dummy in the inefficiency component. However, most of the studies on efficiency effects of merging apply are based on the non-parametric DEA, subject to our attention below.

Annual or global frontiers
In measuring efficiency using panels, most published work computes efficiency scores based on annual frontiers (see Emrouznejad and Yang 2018 for an overview). However, Kjekshus and Hagen (2007) and Papadimitriou and Johnes (2018) are examples that use a pooled frontier for the whole time period as benchmark when measuring the impacts of mergers. These studies incorporate time heterogeneity, *ex post*, by time fixed effects in the second stage. Dynamics related to mergers may also be captured by a Malmquist productivity index (MPI) and its components as performed by, e.g., Ferrier and Valdmanis (2004), Odeck (2008) and Agrell et al. (2015).

---

[6] Silva (2018) does not include judges since each bench within a court is represented by one judge—and no aggregation to court level is performed. In addition, the main objective is to link inputs to outputs for Portuguese courts where the linkages go through cases received and cases solved in different categories.

Identification of impacts

Several empirical strategies are reported in the literature. Harris et al. (2000) use an identification design where merged courts after a merger are compared to a hypothetically merged court before mergers. To construct the hypothetical merged court, Harris et al. (2000) sum the inputs and outputs of the units that are included in the merge, *ex ante*. A problem with this approach is that no control group is used. A second strand of the literature identifies two groups: merged (treated) and non-merged (control) units. To identify the treated and control groups, Ferrier and Valdmanis (2004) use case matching, Schmitt (2017) uses Mahalanobis distances and Dranove and Lindrooth (2003) performed propensity score matching. Examples of matching variables are service provision, along with organizational form and size (Ferrier and Valdmanis, 2004). If a treated and a control group are identified both before and after merger, a DiD approach is possible. Several studies apply DiD related to performance effects of mergers. For example, Kwoka and Pollitt (2010) investigate the impact on technical efficiency and Dranove and Lindrooth (2003), Azevedo and Mateus (2014), and Schmitt (2017) investigate the impact on cost using a DiD approach. Recent work by Bogetoft and Kromann (2018) use a propensity-based matching approach in a DEA setting to create more plausible peers in an applied setting. They show by simulations that the matching approach reduces bias in the frontier estimates, especially for dynamic assessments of frontier shifts.

Timing

Achieving the intended efficiency gain may take time. Several of the previously mentioned studies only investigate the effect during a short period, e.g. Harris et al. (2000) and Ferrier and Valdmanis (2004) use 1 year as the post-treatment period. Worthington (2001) and Groff et al. (2007) use a slightly longer period of at least 2 years, but notes that a longer time period would be preferable. A longer time period is investigated by, for example, Kwoka and Pollitt (2010) and Papadimitriou and Johnes (2018), where potential heterogeneity between the different post-treatment years is captured by time-fixed effects. Regarding *ex post* effects, Kwoka and Pollitt (2010) argue that the most important period is 2–5 years after the merger, since merging effects cannot be exploited the first year. Further, changes later than 5 years afterwards are argued not to be assignable to the merging.

Results

The empirical results of mergers do not point in the same direction and vary between sectors. Two examples where mergers were not considered beneficial are Kjekshus and Hagen (2007) and Azevedo and Mateus (2014). However, Çelen (2013) and Papadimitriou and Johnes (2018) found positive results regarding technical efficiency, and Schmitt (2017) concluded that there were cost savings.[7] Time heterogeneity has also been observed in a few studies. Papadimitriou and Johnes (2018) found that the positive effect on efficiency disappeared after 1 year, and Groff et al. (2007) did not find any effect the first year, but found significantly positive effects in the second year.

In summary, previous literature reveals both valuable insights into the research design but also leaves some open questions regarding the impact of mergers. A first issue relates to the identification of merger impacts, i.e. only a few studies use a design in which merged courts are compared with non-merged units. There is no bulletproof strategy for how such a control group should be constructed, since the merged courts do not exist before the merger and the control group is not merged. Therefore, any impact study of mergers has to rely on assumptions and a perfect identification does not exist. A second problem is when to measure

---

[7] Lynk (1995) and Azevedo and Mateus (2014) found that mergers are more likely to become successful when units of similar size merge. Further, Lynk (1995) concluded a reduction in variability.

the merger effect since the impact may take several years. Our approach to this problem is to use several years of investigation in the follow-up period and to apply different analytical frameworks, i.e. a triangulation of the results.

4 Application

Impact evaluations require relevant units of comparison, i.e. treated and controls, to conduct a reliable estimation of the treatment effect. The first part of the methodology describes how the technical efficiency is measured. Second, a description of how we attempt to identify the treatment effect is provided. Third, the dataset is described.

4.1 Technical efficiency

In a non-parametric framework like DEA (Farrell 1957; Charnes et al. 1978), observations called decision making unit (DMU), defined as a unit i at a given time t, are vector of N inputs, $\mathbf{x}_{i,t}$, to produce a vector M outputs, $\mathbf{y}_{i,t}$. The observations form a technology or production possibility set S defined as

$$S = \left\{ \left(\mathbf{x}_{i,t}, \mathbf{y}_{i,t}\right) \middle| \mathbf{x}_{i,t} \text{can produce } \mathbf{y}_{i,t} \right\}, \tag{1}$$

Technical input-efficiency, TE, is here defined as the maximal radial contraction (or distance measure) of inputs that can be made at constant output level, i.e., a coefficient for each unit between 0 and 1. Formally, TE is obtained as $1/D\left(\mathbf{x}_{i,t}, \mathbf{y}_{i,t}\right)$ from S through

$$D\left(\mathbf{x}_{i,t}, \mathbf{y}_{i,t}\right) = \min\left\{\theta : \left(\theta\mathbf{x}_{i,t}, \mathbf{y}_{i,t}\right) \in S\right\}, \tag{2}$$

where θ is a non-negative scalar. The technology can be estimated as a linear hull using a cross-section for 1 year, or as a pooled frontier using panel data. Estimations of dynamic changes, such as technical change and efficiency change, require the application to an annual frontier as in the Malmquist decomposition. However, since we study the specific impact of mergers independently of the actual year when they occurred, and moreover intend to correct for the number and size of the groups of merged and non-merged DMUs, we do not apply the Malmquist approach.

The radial technical input-efficiency score under constant returns to scale (CRS) is obtained from a linear program calculating the distance $D\left(\mathbf{x}_{i,t}, \mathbf{y}_{i,t}\right) \leq 1$ (Charnes et al. 1978).[8]

$$D\left(\mathbf{x}_{i,t}, \mathbf{y}_{i,t}\right) = \min \ \theta \tag{3}$$

$$\text{s.t.}$$

$$\theta\mathbf{x}_{i,t} \geq \sum_{k=1}^{I} \lambda_{k,t}\mathbf{x}_{k,t} \tag{4}$$

$$y_{i,t} \leq \sum_{k=1}^{I} \lambda_{k,t}y_{k,t} \tag{5}$$

$$\lambda_{k,t} \geq 0 \qquad \text{for } k = 1, 2, \ldots, I \tag{6}$$

---

[8] The rule of thumb suggested by Simar and Wilson (2000) suggests that the bias should not be corrected for unless $s^2 < \frac{1}{3}\left(Bias_B\left[\hat{\theta}(\mathbf{x}_{i,t}, \mathbf{y}_{i,t})\right]\right)^2$, s^2 where represents the variance of the bootstrapped values. In our case, this relation is $0.00016 > 0.00000009$ and the bias is, therefore, not corrected for.

The last restriction can be changed to $\sum_{k=1}^{I} \lambda_{k,t} = 1$ under the variable returns to scale (VRS) assumption (Banker et al. 1984) which is only used for the calculation of scale efficiency.[9] In addition to the computed efficiency, the impact on efficiency of merging is compared to the potential *ex ante* efficiency gain calculated by Mattsson and Tidanå (2019). These results are obtained by applying the Bogetoft and Wang (2005) model where the potential efficiency gain is measured as: overall efficiency gain, i.e. the total estimate of potential gain of merging obtained by summing the initial court´s inputs and outputs. The overall effect is decomposed into learning (technical), harmony (scope), and scale efficiency effects. Learning is an estimate of gain achieved by eliminating the initial inefficiency. Here, we use two components to be compared with the conditional DiD estimates. First, an overall measure as the product of the three components, i.e. overall potential efficiency gain is used. The second estimate excludes effects of becoming fully efficient, which is referred to as learning-adjusted potential efficiency gain.[10]

4.2 Empirical strategy

Identifying the effects of mergers is, in general, problematic. The main issue is the fact that the same units do not exist before and after the merger, making it a challenge to find a unit of comparison of relevance to identify the merging effect. This means that either pre- or post-treatment units have to be constructed. We use three methodological frameworks to address this issue. The first two frameworks are non-standard in policy evaluation and used to support the results found when applying the third framework, which is a conditional DID, i.e. a frequently used method to assess impacts of policy changes.

The first framework relies on the overall objective for the sector. Merging district courts should increase the sector efficiency or reduce its inefficiency. Thus, by studying the aggregate inefficiency between years, a measure of the total inefficiency in the sector is obtained. This is obtained by using input-weighted measures of efficiency. The total inefficiency for each year is therefore expressed as how many working hours of judges, law clerks, other staff and how much office area that can be saved.

In the second framework we make use of the fact that two groups of courts existed after 2009. The groups are courts that have merged and courts that never have merged. The meta-technology is constructed from all observations of merged and non-merged district courts.[11] In addition, group frontiers are constructed based on merged and non-merged courts. Technical efficiency is separated here into managerial efficiency and organizational efficiency (Charnes et al. 1981; Grosskopf and Valdmanis 1987; Månsson 1996).[12] Managerial inefficiency is defined as the amount of inputs that can be reduced compared to peers belonging to

---

[9] The individual courts cannot change their jurisdiction limits, set by SNCA, speaking in favor of a VRS assumption for the assessment of individual efficiency of court managers. However, our focus here is on the contrary at the horizontal restructuring process by SNCA, explicitly changing the profile and size of the courts. This assessment should naturally be based on a comparison with most productive scale at the CRS level to enable scale efficiency analyses.

[10] An alternative to the *ex ante* approach by Bogetoft and Wang (2005) is the method suggested by Wu et al. (2011).

[11] An alternative to use the metafrontier approach is to follow Camanho and Dyson (2006) and compare groups using the components of the Malmquist index.

[12] Much of the recent literature references about metafrontiers are made in relation to O'Donnell et al. (2008). However, metafrontier approaches have been used before (see e.g., Charnes et al. 1981; Grosskopf and Valdmanis 1987; Månsson 1996).

the same group, i.e. merged or non-merged courts. Organizational efficiency is defined as the ratio of the efficiency scores using the pooled frontier over the group frontier, i.e. TE_P/TE_G where the P subscript represents pooled and G denotes the group (merged, non-merged).[13] A hypothetical merging effect from becoming larger in size is that large courts are potentially less sensitive to changes to justice demand. Lower efficiency during some years affects the group average and will be captured in the organizational efficiency component. To test for differences between merged and non-merged courts with respect to organizational and/or managerial efficiency, we apply the non-parametric Mann–Whitney U test.[14]

In the third framework, a conditional DiD is performed. Hypothetical pre-treatment courts are constructed by adding the inputs and outputs for the merged district courts before the actual merging occurred. These observations will be used as pre-treatment units. To obtain better balance between treated and controls, we perform matching to identify twins. This approach allows the construction of more sensible counterfactuals for managerial action (Cobb-Clark and Crossley 2003; King et al. 2017). Matching as an identification strategy is common in, for example, labor economics (see e.g., Lechner 2002). However, it is still not commonly used within the area of efficiency analysis. Several matching methods are available, for example, propensity score matching (Rosenbaum and Rubin 1985), coarsened exact matching (CEM) (Blackwell et al. 2009) and Mahalanobis distance (Mahalanobis 1936; Rubin 1980). An argument for using matching instead of, for example, controlling on observables is that the matching is performed *ex ante*, which identifies a balance before treatment, but allows for differences in outputs *ex post*. A difference in outputs *ex post* is therefore interpreted as the merging effect and identified by the DiD estimate.

This paper adopts Mahalanobis distance which, as with other matching procedures, has pros and cons.[15] To choose relevant matching criteria, we first note that the size will change by construction when merging. An internal argument for mergers within the Swedish district courts was, for example, that recruitment of qualified personnel would be easier at a larger court (Ministry of Justice 2000). Thus, the chosen matching variable focuses on a measure of scale by targeting the *ex ante* sum of each output category as a proxy of size.[16] The court with the closest distance is chosen as the match twin.[17] In the pre-merger (pre-treatment) period we will have one hypothetical group of courts that will be merged the following year and one group of hypothetical courts that do not merge at any point in time. The groups are similar with respect to the sum of outputs during the pre-merging period. The period of investigation is summarized in Fig. 1.

Matching takes place the year just before merger $(t-1)$. The merging year, t, is eliminated from the analysis, since the aim of this study is to evaluate the *ex post* effects of the merger, not

---

[13] For a graphical description see Fig. 9 in the "Appendix".

[14] The Mann–Whitney U test was, in the court context, applied by Santos and Amado (2014) to investigate differences between large and small courts.

[15] Our approach identifies the matches by using the closest Mahalanobis distance. There are alternatives, for example coarsened exact matching (CEM) proposed by Iacus et al. (2012). Using CEM and a one-to-one match restriction involves a random draw of one observation from a coarsened strata. This procedure does not necessarily provide the closest match.

[16] Matching has also been performed targeting both scale and scope, i.e. the output vector, without aggregation. Our results are equivalent with that procedure. However, the differences in means of the descriptive statistics shown in Table 3 become larger if matching is performed on each output variable. Thus, matching is, instead, performed on the sum.

[17] This means that one control court can be the best match for more than one treated unit.
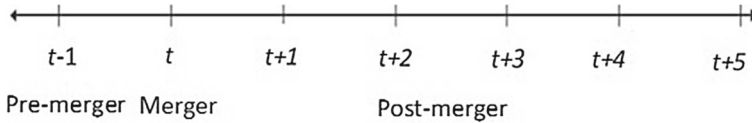
Fig. 1 Time-line of the used data for the conditional DiD approach

the merger-year process.[18] The follow-up period is defined as one $(t+1)$, two $(t+2)$, three $(t+3)$, four $(t+4)$ and five $(t+5)$ years after the respective mergers. The second-stage estimation is performed by the Simar and Wilson (2007) truncated regression approach conducted with the Badunenko and Tauchmann (2016) package.[19] Finally, the merging effect, i.e. the conditional DiD estimates between the merged courts and their matched pair, is individually compared to the potential *ex ante* gain of merging. This procedure makes it possible to evaluate whether the Bogetoft and Wang (2005) model has any predictive power when comparing it to the merging outcome in this application.

4.3 Data

The specification of relevant inputs and outputs is based on several sources. Our model draws on previous research on efficiency and productivity within courts, interviews with representatives from the courts and SNCA and, finally, economic theory. The collected data is an unbalanced panel data set from SNCA that includes all individual district courts for the period 2000–2017 with outputs separated into 14 categories. The model specification includes an aggregation of these categories into three main case categories stated by SNCA: decided civil cases, decided criminal cases, and petitionary matters.[20] A potential problem with aggregation is heterogeneity within the aggregated outputs between courts. This may generate biased efficiency scores if the different categories are not equally distributed between courts. To handle this potential problem, self-reported time consumption is used as approximation of spent resources, and weights for the aggregation are constructed based on these. Self-reported time consumption is used by SNCA for aggregation purposes and generally accepted internally.[21]

On the input side, labor is divided into three categories due to the cost impact (70% of total cost, SNCA 2001–2017)[22] and the different employment conditions, court-differentiated staff composition, and task assignments. These are judges, law clerks, and administrative employees measured as full-time equivalents. As a proxy capital variable, we use office area

---

[18] Observations during the merging year are frequently partially reported, subject to one-shot restructuring effects (offices waiting to be evacuated, vacant staff positions, etcetera) that distort the analysis and the relevance of the results.

[19] Given that the efficiency scores of a merged court and a matched control is $\theta_{i,t}$ and $\theta_{j,t}$ respectively, during time t and similarly during $t+1$, we can formulate the DiD estimate (the treatment effect) as $(\theta_{i,t+1} - \theta_{j,t+1}) - (\theta_{i,t} - \theta_{j,t})$. This is identified using the equation $\theta_{i,t} = \beta_0 + \beta_1 Treated + \beta_2 Post\text{-}merger + \beta_3 DID + \varepsilon_{it}$ (see e.g., Card and Krueger 1994; Angrist and Pischke 2008).

[20] Property cases and environmental cases are included in the category criminal cases to avoid many courts with one output equal to zero. This is a reasonable procedure according to SNCA.

[21] For example, within criminal cases, both summary offences and environmental cases are aggregated. However, the average time consumption for handling an environmental case is 5.7 times larger than the average time for a summary offence. Therefore, an environmental case receives the weight of 5.7 compared to a summary offence.

[22] The labour cost is approximately 70% of total cost during each year of the studied period except 2000–2001 when it is 60%, as extracted from SNCA annual reports 2001–2017 (from webpage www.domstol.se, in Swedish).

Table 2 Descriptive statistics, separated by mergers and non-mergers, for 2000–2017	All $n = 997$	Merged $n = 655$	Non-merged $n = 342$
Inputs			
Office area	3629.31	4227.66	2483.36
Judges	11.31	13.17	7.74
Law clerks	10.88	12.87	7.07
Other staff	21.48	24.87	14.99
Outputs			
Criminal cases	1465.70	1723.17	972.59
Civil cases	1036.45	1221.83	681.41
Petitionary matters	475.97	550.73	332.79

*The weighted means are reported as the outputs

with the assumption that the size of the premises is proportional to other capital variables, for example, the number of computers and other equipment, but also operational expenditure such as heating, maintenance, and insurance.[23]

Regarding quality, it can be argued that there is a trade-off with efficiency. Thus, leaving quality variables out of the model would potentially give biased results, prompting discussions within and outside of the court system. Explicitly, the question of whether the differences between district courts can be observed in terms of the rate of change in higher instances has been raised. In an attempt to respond to this question, Andersson et al. (2017), using a subset of our data, included a correlation analysis between rates of change in higher instances and efficiency. The result pointed to a non-significant correlation of -0.13. The low correlation indicates quality differences in this aspect cannot be observed. Furthermore, cases with new evidence, which is one source of changed decisions in a higher instance, are likely to be distributed evenly between district courts.[24]

Descriptive statistics of the average level of inputs and weighted outputs for the courts included in the mergers and courts not included in mergers, respectively, are reported in Table 2.[25]

4.3.1 Matching quality

As described in the final part of the identification strategy, a conditional DiD will be conducted. In order to obtain comparable groups, matching was performed based on the sum of outputs. Descriptive statistics of $t-1$, $t+1$, $t+2$, $t+3$, $t+4$ and $t+5$, i.e. the time periods

---

[23] Under varying local input prices, a cost efficiency analysis could also be relevant in investigating the allocative efficiency with respect to substitution between labor and capital inputs. However, as main capital input (the court buildings) can be considered asset and location specific, e.g. a court cannot receive the proceeds from renting out or selling the asset to offset staff cost. Thus, the cost approach is not relevant for our application because also the staff salaries are equal according to SNCA. Hence, our approach considers only technical efficiency.

[24] A potential issue in relation to merging district courts is that the mergers experience a higher degree of congestion that can generate more delays. Data on non-weighted caseload, defined as the stock in the beginning of the year plus incoming cases during the year, is available for the period 2011–2015. No statistically significant difference in percentage change in caseload can be observed between merged and non-merged courts. This also holds measured as a difference between one year and the next.

[25] A back-up force of judges served courts in the Stockholm region during the period 2008–2010. Similarly, it served nationally during the period 2013–2017. Each court that used personnel from the back-up force have been assigned these hours equivalent to the time spent.

Table 3 Descriptive statistics for the difference-in-difference estimation

	(1) All non-mergers	(2) Matched non-mergers	(3) Matched mergers	(4) Difference non-matched	(5) Difference matched
Inputs					
Office area	2928.9 (1994.1)	4639.1 (2021.8)	4078.7 (2607.0)	1149.8***	− 560.4
Judges	8.656 (6.936)	11.49 (6.286)	11.93 (7.874)	3.274***	0.44
Law clerks	8.228 (5.755)	12.44 (6.587)	11.60 (8.072)	3.372***	− 0.84
Other staff	17.79 (14.78)	24.89 (13.92)	22.77 (17.19)	4.98***	− 2.12
Outputs					
Criminal cases	1132.8 (866.3)	1517.7 (920.1)	1619.7 (1010.2)	486.9***	102.0
Civil cases	778.5 (676.5)	1058.5 (701.1)	1086.2 (971.0)	307.7***	27.7
Petitionary matters	443.0 (285.5)	647.9 (275.1)	532.6 (333.8)	89.6***	− 115.3**
	779	167	167		

*The weighted means are reported as the outputs. Standard deviations are shown in parentheses
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

to be used in the final analysis presented in Sect. 5.3 are aggregated and reported in Table 3 separated on the matched and non-matched data.

In column 1 of Table 3, non-mergers include both the hypothetical courts created to be compared with the mergers and non-merged courts that were not matched. Column 2 and 3 eliminate courts that neither merged nor became matched with a merge.[26] Comparing column 1 and 3, it can be observed that the mean values in column 1 are smaller—this difference is shown in column 4 and is significant on the 1% level in each case. Comparing column 2 and 3, the matched groups are more similar than previously and column 5 shows that the differences after matching are non-significant on each input and output with the exception of petitionary matters. Therefore, regardless of whether the matching was performed on outputs, the differences in inputs are also reduced. The remaining difference for petitionary matters can be considered problematic. However, this is handled by weighting the outputs based on complexity within different categories, because no single matching method will be able to fix all differences in the data (Iacus et al. 2012). In addition to similarities in characteristics, one assumption in the DiD frameworks is pre-treatment parallel trends. Therefore, a visual analysis of efficiency scores over time is presented in Fig. 2.

Figure 2 shows the trends over time for the full and matched samples to be used in the DiD and conditional DiD estimations, respectively. Both the full and matched samples indicate that the efficiency of the merged courts declined during the pre-treatment period, i.e. $t − 2$ and $t − 1$. However, a trend downwards is not as much of an issue as the other way around. For example, a trend upwards *ex ante* would indicate that these courts would be about to perform better, with a different trend, and also without the merger. A potential source of the decline of the efficiency score for merged courts, *ex ante*, is preparation for the merger.[27]

---

[26] The column to represent all mergers is eliminated since it is equivalent to matched mergers. Further, courts that are merged more than once are eliminated in that post-merger period and the following, e.g. if merged again in $t + 4$, that court is eliminated from the post-treatment group during $t + 4$ and $t + 5$.

[27] Selection bias is a common threat in all quasi-experimental frameworks, e.g. DiD. We attempt to minimize the selection bias by matching on the main selection criteria, size, *ex ante* and thereafter perform a DiD.
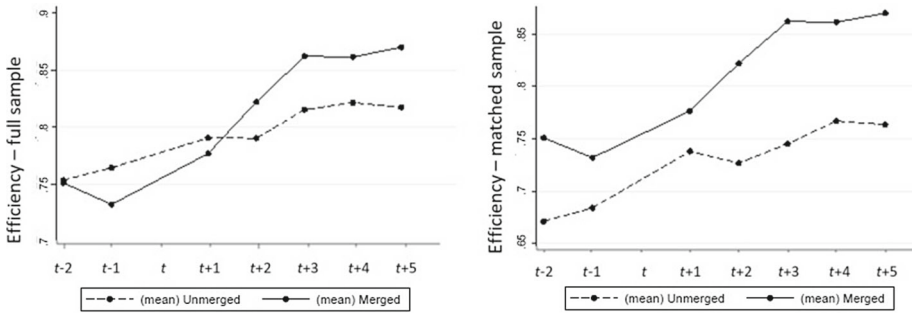
Fig. 2 Visual analysis of the pre-treatment trends using the full and matched samples
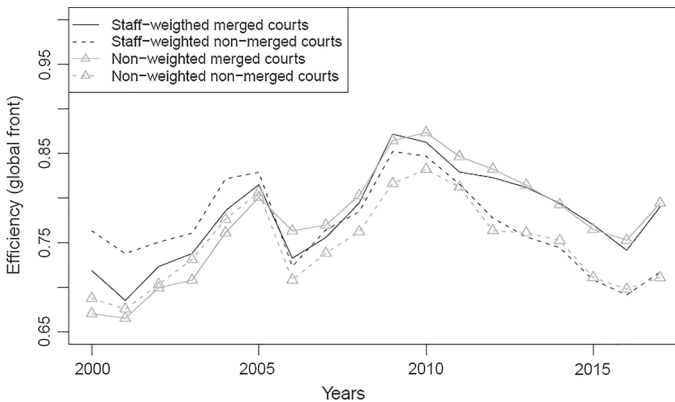

Fig. 3 Efficiency scores (global frontier) over time under CRS

5 Results

The empirical analysis is presented from our three analytical frameworks: (1) an overview of how the average performance has changed over time is provided together with sources of the development, (2) a metafrontier analysis is performed over the time period 2012–2017 separated between merged and unmerged courts, and (3) a conditional DiD procedure of merged courts in relation to their matched twin is presented followed by an analysis of the sources of efficiency change.

5.1 Global frontier over time

In Fig. 3 we present efficiency scores over time under CRS using a global frontier. The graphs are separated on staff-weighted averages (lines) and arithmetic means (triangles) divided into merged (solid lines) and non-merged (dashed lines) courts.[28]

Figure 3 shows that the staff-weighted average efficiency scores, in relation to the global frontier, are higher for non-merged courts at the beginning of the period by comparing the solid and dashed lines. Furthermore, all efficiency indicators increased fairly constantly for

---

[28] Staff-weighted efficiency scores are obtained as a weighted mean based on the size of the courts measured as number of employees, i.e. the sum of the three staff-categories.
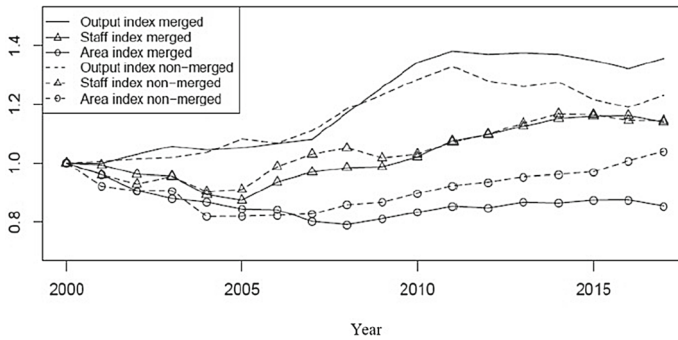
Fig. 4 Input and output development (index 2000)

all courts during the period 2000–2010. The exception is a decline in 2005–2006. This initial period of witnessing efficiency increases corresponds to the period during which the mergers took place. This suggests that merging was advantageous from a short-term efficiency point of view. A statistical examination of the efficiency scores is performed by a regression on time fixed effects. The results reveal that efficiency was significantly larger each year after 2003 compared to the year 2000.[29] The increase until 2010 is followed by a decrease until 2016, a period that may be considered as the post-merger follow-up period. Weighted and non-weighted, merged courts have lower efficiency than the non-merged at the beginning of the period. Further, the lines are crossing in the middle of the period and the non-merged courts are approximately 5 percentage points lower after 2011, regardless of whether the weighted or non-weighted means are used as measure of comparison. This gives an indication of a positive merging effect since the poorest performing courts at the beginning of the period are merged, and at the end of the period, this group performs better, on average, in comparison to the non-merged courts.[30] Sources of these changes are graphically presented in Fig. 4.[31]

Figure 4 shows the development of the staff index represented by the sum of the full-time equivalent, the area index and an output index represented by the sum of the (weighted) decided criminal cases, civil cases and petitionary matters, separated on merged (solid lines) and non-merged (dashed lines) courts. Differences can be observed for outputs (normal lines) which become larger for the merged group after 2007, driven by a larger increase until 2011, and a decrease of non-merged courts after 2011. Furthermore, the office size decreases for both the merged and non-merged courts until 2007, which is indicated by the circle lines. During the following period, i.e. 2008–2017, the merged courts remain at similar or slightly higher levels, while non-merged courts are at a similar level as in the beginning of the period. The percentage increase of office area is in magnitude 25%, i.e. from 80% of its initial level to 100%, regardless of whether the number of employees increased similarly between the two groups.

---

[29] These results are reported in Table 11 in the "Appendix".

[30] In addition, the scale component measured as the ratio between the CRS efficiency scores over the VRS efficiency scores is shown in Fig. 6 in the "Appendix". The arithmetic mean, by year, is observed to decline slightly over time. However, no differences can be observed when weighting the scale component based on size.

[31] Exact correspondence with the efficiency scores is not observed since courts that merged in a specific year are eliminated from the benchmark that year. However, all courts are included in the output and input indexes since a high volatility would occur if courts that merged a specific year were excluded.

Table 4 Efficiency scores based on the metafrontier

	Mean	Min	1st Qu.	Median	3rd Qu.	Max	SD
Overall input efficiency							
Pooled	0.827	0.545	0.761	0.824	0.899	1.000	0.097
Managerial efficiency							
Merged	0.884	0.700	0.828	0.878	0.945	1.000	0.076
Non-merged	0.863	0.582	0.793	0.871	0.948	1.000	0.103
Organizational efficiency							
Merged	0.965	0.879	0.943	0.970	0.987	1.000	0.028
Non-merged	0.913	0.665	0.892	0.922	0.956	1.000	0.063

Table 5 Mann-Whitney U test

	Managerial efficiency z (prob z)	Organizational efficiency z (prob z)
Merged versus non-merged courts	$-$ 1.002 (0.158)	$-$ 8.158 (0.000)

In summary, the first analytical framework points in the direction that merged courts perform better than non-merged. A larger decrease in the office area and higher increase in outputs are indicated to be the sources.

5.2 Metafrontier approach

The metafrontier are created by separating the treated and the untreated district courts into single frontiers to compare the performance of merged and non-merged district courts. All single units during the time period 2012–2017 represent the meta-technology. The sub-frontiers are represented by the merged and non-merged groups. The results are reported in Table 4.

In Table 4, the mean of the efficiency scores during the period 2012–2017 (when no mergers took place) is 0.827 using a global frontier. Using the group as reference technology, merged courts are represented by district courts present during 2012–2017 that have undergone, at least, one merge during the period 2000–2009. These results show that the managerial efficiency for merged courts (non-merged) is 0.884 (0.863). Furthermore, the organizational efficiency is 0.965 and 0.913 for merged and non-merged courts, respectively. This means that both components are larger for the merged group, on average. To investigate whether these differences are statistically significant, Table 5 reports the results of the Mann–Whitney U test.

As shown in Table 5, the differences in managerial efficiency are non-significant according to the Mann–Whitney test, i.e. in relation to its own reference technology, merged and non-merged efficiencies do not differ. In contrast, organizational efficiency scores are significantly different at the 1% level, i.e. the merged courts are organizationally more efficient than the non-merged courts during the period of investigation. This strengthens the arguments that the merging, on average, is advantageous for efficiency.[32] Furthermore, the courts within the

---

[32] The Simar and Wilson (2007) package in STATA (Badunenko and Tauchmann 2016) automatically eliminates efficient courts from the number of observations. Further, a few courts merge again during the post-treatment period and are therefore eliminated from that year and the following one.

Table 6 Difference-in-differences estimations for the matched sample (The Simar and Wilson (2007) package in STATA (Badunenko and Tauchmann 2016) automatically eliminates efficient courts from the number of observations. Further, a few courts merge again during the post-treatment period and are therefore eliminated from that year and the following one)

Variables	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
Treated	0.038	0.037	0.037	0.037	0.038
	(0.028)	(0.026)	(0.023)	(0.025)	(0.028)
Post-merger	0.021	0.045*	0.062***	0.084***	0.071**
	(0.027)	(0.026)	(0.024)	(0.026)	(0.030)
Conditional DiD	0.037	0.057	0.078**	0.041	0.081*
	(0.040)	(0.037)	(0.036)	(0.039)	(0.046)
Constant	0.688***	0.688***	0.688***	0.688***	0.688***
	(0.020)	(0.018)	(0.016)	(0.018)	(0.020)
Observations	117	116	113	107	105

Standard errors in parentheses, ***$p<0.01$; **$p<0.05$; *$p<0.1$

merged group are larger than the non-merged courts, meaning that positive effect of size is another potential cause.

The first and second part of the analytical framework point in the direction that merged courts are more efficient than non-merged. However, neither of these approaches is a proper method for policy evaluation. Therefore, the final part of the analysis adopts a conditional DiD in an attempt to investigate the merger impact.
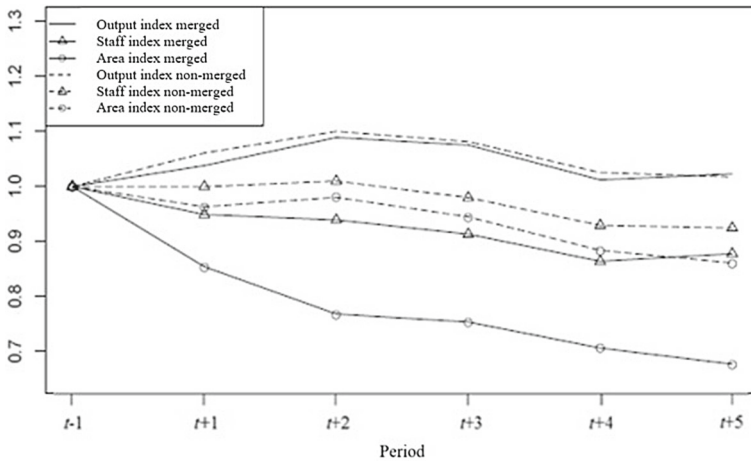
5.3 Conditional DiD

The conditional DiD approach is applied to the matched sample using different post-treatment periods reported in Table 6.

In Table 6, each column represents different time periods post-treatment, i.e. $t+3$ only includes the year three after the merger as the post-treatment period. In the first row of Table 6, it can be observed that no significant difference is present between merged courts and matched non-merged courts. Furthermore, district courts are, on average, more efficient during the post-treatment period observed from the second row, i.e. after. The coefficient of interest is, however, the conditional DiD, which identifies the difference between the merged courts and the hypothetical mergers before in comparison to after the merger took place, i.e. it attempts to represent the treatment effect. The conditional DiD coefficient shows a positive sign during each time period after the merger. Heterogeneity can, however, be observed regarding time because differences are observed with respect to significance and magnitude. The largest magnitude can be observed for $t+3$ and the smallest difference is for $t+1$. Time heterogeneity is not surprising since non-controllable changes of the courts may also occur, e.g. because of volatility in the caseload. The non-matched data indicate higher magnitude and stronger significance of the merging effect.[33]

---

[33] Table 8 in the "Appendix" reports the results for the full sample. These results indicate a stronger significance for the DiD estimator, i.e. significant during each year except $t+1$. Furthermore, the magnitude ranges from 5.0 to 9.5%.

Fig. 5 Matched sample of merged and non-merged courts during $t-1$ to $t+5$ (index $t-1$)

In order to identify the sources of the changes, the development of input and output indexes are reported over the time period studied in the conditional DiD part. The indexes are normalized in period $t-1$ and reported in Fig. 5.[34]

Figure 5 shows that the output indexes develops fairly similar for the non-merged courts and the merged courts symbolized by the dashed and solid lines, respectively. In contrast, the staff index indicates a larger decrease of employees within the merged group, in comparison to the non-merged. An initial reason for this can be that all employees did not move to the new court house. Further, the size of the premises for the merged courts declined steadily after merging. A sharp and direct decrease is not observed because a few merged courts have rental contract, meaning that it takes time before a full reduction in office area can be observed. Office area declines for the control group over time, but more so for the merged courts, i.e. the control group has approximately 90% (circled dashed line) of its initial office space in the end of the follow-up period to be compared with around 70% for the mergers (circled solid line). This corresponds to Fig. 3, where it could be observed that the total office area of the non-merged courts increased. Hence, it suggests that merging led to a higher degree of area efficiency. However, it may be questioned whether this is a pure merging effect, i.e. a scale or scope effect, or if this improvement could have also been obtained without merging. For example, a merger may create the need for new office space and while obtaining the new premises, more thought might have been given to optimal office area utilization, i.e. the dimensioning and allocation of facilities and equipment to tasks and staff. This indicates that there was an office area inefficiency to be eliminated for these courts, *ex ante*. On the other hand, it can be interpreted as a merging effect if larger size prevent empty spaces, e.g. smaller courts may not be able to use their hearing rooms to the same extent as the larger.

Finally, an advantage with the performed matching procedure is the possibility of evaluating each individual merger in comparison with its respective match. These results, combined with *ex ante* potential efficiency gains under non-decreasing returns to scale (NDRS) from Mattsson and Tidanå (2019) are graphically reported in Fig. 6. The latter effects are decom-

---

[34] One robustness test that has been performed is to match on the mean outputs of $t-1$ and $t-2$. Further, the pre-treatment period is then defined as the mean value of inputs and outputs during these 2 years. The results on the merging effect remain consistent.
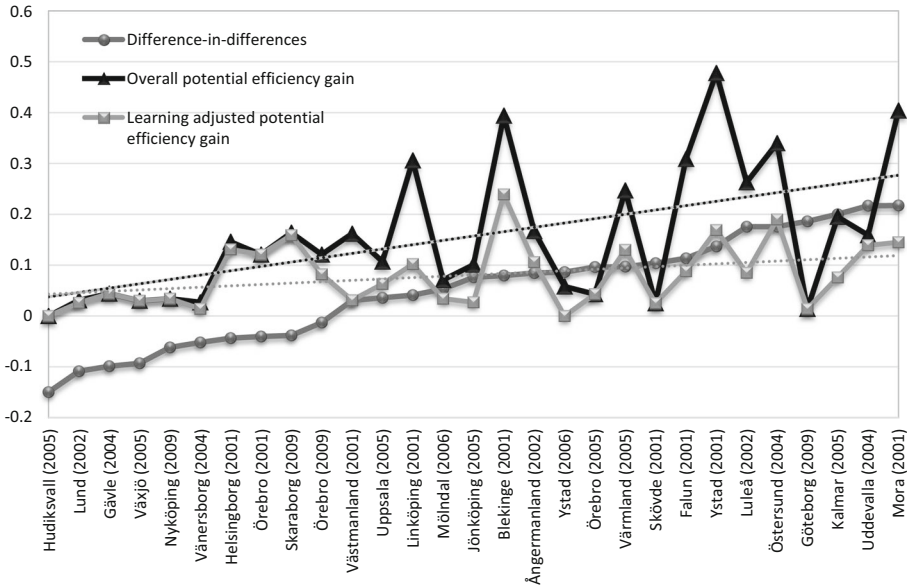
Fig. 6 Individual merging DiD calculated as $t-1$ in comparison to the average of the post-merging periods

posed in an overall potential efficiency gain (black line) and the learning-adjusted potential efficiency gain (squared gray line), respectively.

In Fig. 6, the DiD are calculated separately for each merger, as the difference between $t-1$ and an average over the whole post-treatment period ($t+1$ to $t+5$), in comparison to its matched twin symbolized by light gray circle line. These are ordered from the most negative on the left to the most positive on the right. Averages (standard deviations) are 0.052 (0.105), 0.157 (0.133) and 0.081 (0.063) for DiD, overall potential efficiency gain and learning-adjusted potential efficiency gain, respectively. The Spearman correlation between the DiD and the overall potential efficiency gain is 0.528 with p value 0.003 (the dashed black line). Similarly, the learning-adjusted efficiency gain has a correlation with the DiD of 0.367 with p-value 0.050. These components are statistically significant at the 1% and 10% levels, respectively. This means that the model by Bogetoft and Wang (2005) gives significant insights into the merging outcome in our application. Furthermore, the fact that the dashed black line is steeper than the dashed gray indicates that learning, i.e. adoption to the initial frontier, was, at least, partly achieved. A potential mechanism behind this could be assignment of staff from more efficient courts to less efficient merger targets.

In summary, the realized effects of the individual mergers are heterogeneous. Out of 29 studied mergers, 10 (34.5%) district courts showed a gain of less than 10%, 9 (31.0%) experienced a gain of more than 10%, and 10 (34.5%) of the courts suffered a loss in comparison to their matches.

6 Conclusion and policy discussion

In this paper we provide an applied example of a more general approach to the *ex post* evaluation of structural change in public service provision. Our analysis is directed toward

a merger wave within Swedish district courts in 2000–2009, which was triggered by policy concerns relating to increased efficiency. To get a rich impact assessment, we deploy three methodological frameworks. First, efficiency scores obtained from estimation of a pooled frontier is compared over time. Second, a metafrontier approach is performed by comparing the *ex post* efficiency of the merged and non-merged groups in terms of organizational and managerial efficiency. Differences between the groups are tested using the Mann–Whitney *U* test. Third, conditional DiD is applied to evaluate the impacts of merging by comparing the actual mergers in relation to the non-merged twin, i.e. the court with the smallest Mahalanobis distance. Due to the one-to-one matching, individual estimates of the merger impact can be obtained. These results are compared to the potential *ex ante* efficiency gain of merging obtained by Mattsson and Tidanå (2019), who applied the Bogetoft and Wang (2005) model.

Our results show that efficiency is indeed at a higher level by the end of the period, i.e. in 2017 in comparison to 2000. Furthermore, courts that merged at least once during the period 2000–2009 had a similar average efficiency before merging, but were higher from 2006 and onwards, measured as an arithmetic mean. If weighting is performed on size, merged courts has a lower efficiency score in the end of the period (before the mergers took place) and by the end of the studied period, merged courts are better off in terms of efficiency. Our metafrontier shows that the merged group is more efficient than the non-merged. This is driven by a significantly higher organizational efficiency. Finally, the conditional DiD resulted in a positive merging effect during each post-treatment period of a magnitude of between 4.1 and 8.1 percentage points higher efficiency, on average, compared to the matched control group. This effect is statistically significant during three out of five post-treatment years. As a final part, we found a positive and statistically significant correlation between the resulted DiD estimates of the individual mergers and the estimates of potential, *ex ante*, efficiency gain. A caveat must be noted in interpreting the results for initially efficient merged courts (matched controls), since the conditional DiD estimates here are bounded upwards (downwards) to zero.

We conclude that merging the district courts, on average, made the whole sector more efficient. This finding holds when measuring weighted or non-weighted efficiency over time, separating the groups using a metafrontier approach, and when DiD is applied to both the matched and non-matched sample. However, at an individual level, each merger is not better off than its matched pair. Furthermore, our graphical evaluation of the sources of efficiency change showed that one of the main differences between the merged and non-merged courts is that merged courts decrease their office area more, which took several years due to rental contracts. Based on this, a question of whether a change in a sticky (fixed or semi-fixed) input in the public sector is a merger effect that can be raised. For example, it is natural that a merged court would rearrange, acquire, or rent new facilities to cater the new scale of operations. In contrast, it is less likely that an existing court would say that their court house should be fully or partially reallocated for other use or rented out, even if the case load would decrease or call for less staff. Likewise, given the staff employment conditions in the public sector, it is uncommon for an internal productivity improvement project to trigger actual reductions in the permanent staff count, especially for higher civil servants as judges. This is an example of a 'sticky' input, where an increase (creation of a new unit adjusted to expected output) is easy, but a decrease (downsizing of staff or office area) is difficult or impossible. To summarize, mergers give incentives to consider issues of modifying such fixed, or semi-fixed, inputs that are less likely to be changed regardless of their potentials. However, it is also important to take into consideration the social cost of restructuration of public service, both in terms of geographical and social proximity to the users, as well as the impact on existing staff and managers. Therefore, care needs to be taken during the *ex ante* process in examining whether a merger should be performed. To compare with other studies
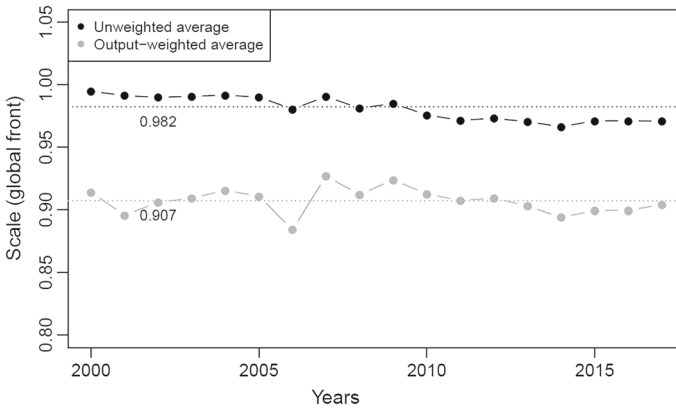
Fig. 7 Average estimated scale efficiency over time using a global front

that investigate how efficiency could be improved in courts from other kinds of reforms, Falavigna et al. (2018) concludes that a reduction of the number of sections had a negative impact on performance but efficiency can be enhanced by using the judges more efficiently. Schneider (2005) and Deyneli (2012) also to put forward that it is possible to implement incentives such as higher salaries for judges.

This paper addressed several relevant policy and research questions, with methodological generality beyond the country and area of application. The sticky characteristic of inputs in public services is common across areas, as are the tendencies to undertake only short-term or *ex ante* assessment of horizontal restructuring. Thus, we believe that the presented approach can be useful in assessing mergers also in other settings. In addition, the conclusion that the Bogetoft and Wang (2005) model has predictive power in our application gives an indication that it would be useful to strengthen the decision support before merging decisions are made. Future research can estimate the social cost of mergers and make a comparison to the achieved gain in order to investigate whether the merging was necessary from a cost–benefit point of view. In addition, it would be necessary to evaluate the mergers qualitatively, e.g. whether merged entities can recruit managers with higher skills.

Appendix

See Tables 7, 8, 9, 10, 11 and Figs. 7, 8, 9.

Figure 9 illustrates how the decomposition can be performed using a metafrontier. The G frontier represents the group, i.e. merged or non-merged courts and the P frontier represents the pooled. The court to be evaluated in this illustration is A which we assume is a

Table 7 Date of the mergers

New court	Merged courts	Date
Sollentuna	Sollentuna and Jakobsberg	July 1, 2000
Hudiksvall	Hudiksvall and Ljusdal	October 1, 2000
Västmanland	Köping, Sala, and Västerås	April 1, 2001
Linköping	Linköping, Mjölby, and Motala	May 1, 2001
Örebro	Örebro and Hallsberg	June 1, 2001
Blekinge	Karlskrona, Karlshamn, Ronneby, and Sölvesborg	July 1, 2001
Falun	Falun, Hedemora, and Ludvika	September 1, 2001
Mora	Mora and Leksand	September 1, 2001
Skövde	Skövde and Falköping	September 17, 2001
Helsingborg	Helsingborg, Klippan, and Ängelholm	October 1, 2001
Ystad	Simrishamn and Ystad	November 26, 2001
Luleå	Boden, Luleå, and Piteå	January 28, 2002
Ångermanland	Härnösand, Sollefteå, and Örnsköldsvik	February 25, 2002
Lund	Eslöv, Landskrona, and Lund	April 15, 2002
Gävle	Gävle and Sandviken	April 26, 2004
Östersund	Sveg and Östersund	May 17, 2004
Uddevalla	Strömstad and Uddevalla	September 20, 2004
Vänersborg	Trollhättan and Vänersborg	October 4, 2004
Kalmar	Kalmar, Oskarshamn, and Västervik	January 17, 2005
Värmland	Arvika, Karlstad, Kristinehamn, and Sunne	February 7, 2005
Örebro	Lindesberg and Örebro	February 14, 2005
Hudiksvall	Bollnäs and Hudiksvall	June 1, 2005
Jönköping	Jönköping and Värnamo	September 1, 2005
Växjö	Ljungby and Växjö	September 12, 2005
Uppsala	Enköping, Tierp, and Uppsala	December 1, 2005
Ystad	Trelleborg and Ystad	January 1, 2006
Mölndal	Mölndal and Stenungsund	December 1, 2006
Attunda	Sollentuna and parts of Södra Roslagen	April 1, 2007
Södertörn	Parts of Handen, Huddinge, and parts of Stockholm	April 1, 2007
Nacka	Nacka, parts of Handen, and parts of Stockholm	April 1, 2007
Solna	Parts of Stockholm and parts of Södra Roslagen	April 1, 2007
Stockholm	Solna and parts of Stockholm	April 1, 2007
Skaraborg	Lidköping, Mariestad, and Skövde	January 12, 2009
Örebro	Karlskoga and Örebro	January 12, 2009
Nyköping	Katrineholm and Nyköping	April 27, 2009
Gothenburg	Gothenburg and Mölndal	October 19, 2009

Table 8 Difference-in-differences estimations for the full sample

Variables	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
Treated	− 0.035	− 0.035	− 0.035	− 0.035	− 0.035
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Post-merger	0.009	0.034**	0.058***	0.062***	0.061***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)
DiD	0.050	0.074**	0.094***	0.070*	0.095**
	(0.033)	(0.035)	(0.036)	(0.037)	(0.041)
Constant	0.761***	0.761***	0.761***	0.761***	0.762***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	307	312	306	298	299

Standard errors in parentheses, ***$p<0.01$; **$p<0.05$; *$p<0.1$

Table 9 Difference-in-differences estimations for the full sample matched on $t-2$ and $t-1$

Variables	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
Treated	− 0.034	− 0.034	− 0.034	− 0.034	− 0.034
	(0.023)	(0.023)	(0.022)	(0.023)	(0.024)
Post-merger	0.010	0.035**	0.059***	0.063***	0.062***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)
Conditional DiD	0.049	0.073**	0.093***	0.068*	0.094**
	(0.033)	(0.035)	(0.036)	(0.037)	(0.040)
Constant	0.761***	0.761***	0.760***	0.760***	0.761***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	307	312	306	298	299

Standard errors in parentheses, ***$p<0.01$; **$p<0.05$; *$p<0.1$

**Table 10** Difference-in-differences estimations for the matched sample matched on $t-2$ and $t-1$

Variables	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
Treated	0.040	0.040	0.040*	0.040	0.040
	(0.028)	(0.026)	(0.024)	(0.024)	(0.027)
After	0.026	0.049*	0.064***	0.091***	0.077***
	(0.028)	(0.025)	(0.024)	(0.027)	(0.029)
Conditional DiD	0.033	0.054	0.076**	0.034	0.074*
	(0.041)	(0.037)	(0.036)	(0.039)	(0.044)
Constant	0.686***	0.686***	0.686***	0.686***	0.686***
	(0.019)	(0.018)	(0.017)	(0.017)	(0.019)
Observations	117	116	113	107	105

Standard errors in parentheses, ***$p<0.01$; **$p<0.05$; *$p<0.1$

Table 11 Regressions of global efficiency scores with time fixed effects

	Model 1	Model 1
(Intercept)	0.67 (0.01)***	0.65 (0.01)***
t_2001	− 0.01 (0.02)	− 0.00 (0.02)
t_2002	0.03 (0.02)*	0.03 (0.02)
t_2003	0.04 (0.02)***	0.04 (0.02)**
t_2004	0.09 (0.02)***	0.09 (0.02)***
t_2005	0.13 (0.02)***	0.13 (0.02)***
t_2006	0.07 (0.02)***	0.07 (0.02)***
t_2007	0.08 (0.02)***	0.09 (0.02)***
t_2008	0.11 (0.02)***	0.12 (0.02)***
t_2009	0.17 (0.02)***	0.18 (0.02)***
t_20010	0.18 (0.02)***	0.19 (0.02)***
t_20011	0.16 (0.02)***	0.16 (0.02)***
t_20012	0.13 (0.02)***	0.14 (0.02)***
t_20013	0.12 (0.02)***	0.12 (0.02)***
t_20014	0.10 (0.02)***	0.11 (0.02)***
t_20015	0.07 (0.02)***	0.07 (0.02)***
t_20016	0.06 (0.02)**	0.06 (0.02)***
t_20017	0.09 (0.02)***	0.09 (0.02)***
Merged		0.03 (0.01)***
R^2	0.24	0.25
Adj. R^2	0.22	0.24
Num. obs.	997	997
RMSE	0.10	0.10

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$
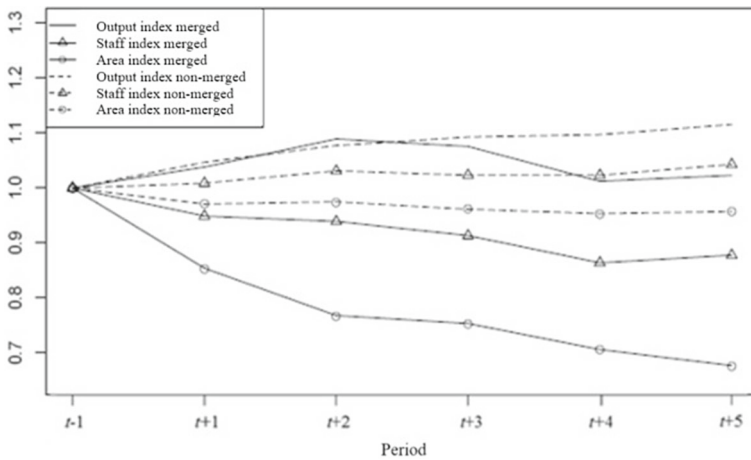
Fig. 8 Full sample of merged and non-merged courts during $t − 1$ to $t + 5$ (index $t − 1$)
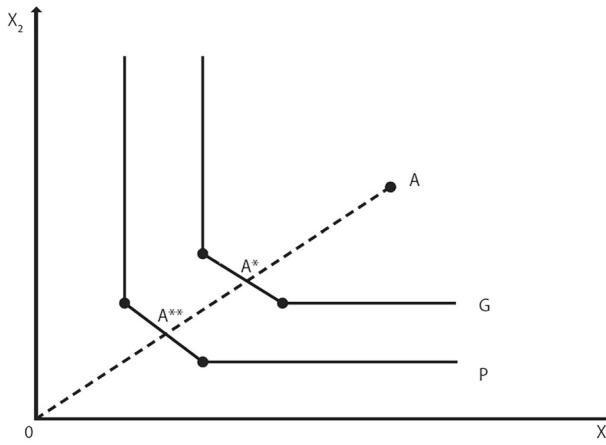
Fig. 9 Illustration of decomposition of efficiency terms in metafrontier

merged court. An input-based framework is used in the analysis where X_1 and X_2 represent two different inputs. The total technical efficiency score for unit A is $\|0A^{**}\|/\|0A\|$. This can be decomposed into two parts, i.e. managerial efficiency and organizational efficiency. Managerial efficiency is computed as $\|0A^*\|/\|0A\|$, i.e. the distance to the group frontier. Furthermore, there is also one part, $\|0A^{**}\|/\|0A^*\|$, i.e. the distance between the group frontier and the pooled frontier. This part cannot be realized by court A, in Figure 9, due to the fact that A belongs to group G. That ratio is labeled organizational inefficiency.

References

Agrell, P. J., Bogetoft, P., & Grammeltvedt, T. E. (2015, May). The efficiency of the regulation for horizontal mergers among electricity distribution operators in Norway. In *12th international conference on the European energy market (EEM)*, 2015 (pp. 1–5). IEEE.

Andersson, C., Bonander, F., & Månsson, J. (2017). *Tingsrätters effektivitet och produktivitet (The efficiency and productivity of Swedish district courts)*. Stockholm: Swedish National Audit Office. **(in Swedish)**.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.

Avkiran, N. K. (1999). The evidence on efficiency gains: The role of mergers and the benefits to the public. *Journal of Banking & Finance, 23*(7), 991–1013.

Azevedo, H., & Mateus, C. (2014). Cost effects of hospital mergers in Portugal. *The European Journal of Health Economics, 15*(9), 999–1010.

Badunenko, O., & Tauchmann, H. (2016). SIMARWILSON: Stata module to perform Simar & Wilson (2007) efficiency analysis. Statistical Software Components S458156, Boston College Department of Economics, revised 31 Mar 2018.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis, 3*(1–2), 153–169.

Blackwell, M., Iacus, S. M., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *The Stata Journal, 9*(4), 524–546.

Bogetoft, P., & Kromann, L. (2018). Evaluating treatment effects using data envelopment analysis on matched samples: An analysis of electronic information sharing and firm performance. *European Journal of Operational Research, 270*(1), 302–313.

Bogetoft, P., & Wang, D. (2005). Estimating the potential gains from mergers. *Journal of Productivity Analysis, 23*(2), 145–171.

Camanho, A. S., & Dyson, R. G. (2006). Data envelopment analysis and Malmquist indices for measuring group performance. *Journal of Productivity Analysis, 26,* 35–49.

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review, 84,* 772–793.

Cartwright, S., & Cooper, C. L. (1993). The role of culture compatibility in successful organizational marriage. *The Academy of Management Executive, 7*(2), 57–70.

Çelen, A. (2013). The effect of merger and consolidation activities on the efficiency of electricity distribution regions in Turkey. *Energy Policy, 59,* 674–682.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.

Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science, 27*(6), 668–697.

Cobb-Clark, D. A., & Crossley, T. (2003). Econometrics for evaluations: An introduction to recent developments. *Economic Record, 79*(247), 491–511.

Deyneli, F. (2012). Analysis of relationship between efficiency of justice services and salaries of judges with two-stage DEA method. *European Journal of Law and Economics, 34*(3), 477–493.

Dranove, D., & Lindrooth, R. (2003). Hospital consolidation and costs: Another look at the evidence. *Journal of Health Economics, 22*(6), 983–997.

Elbialy, N., & García-Rubio, M. A. (2011). *Assessing judicial efficiency of Egyptian first instance courts: A DEA analysis* (No. 19-2011). Joint discussion paper series in economics.

Eliasson, D., Perklev, A., Holmgren, M., Öberg, N., & Wennerström, E. (2017). *Den framtida verksamhetsvolymen i rättskedjan—centrala prognoser 2017–2020* (*The future volume of matters in the legal chain—central forecasts 2017–2020*). Swedish National Council for Crime Prevention. Norrköping. (**in Swedish**).

Emrouznejad, A., & Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences, 61,* 4–8.

Falavigna, G., Ippoliti, R., Manello, A., & Ramello, G. B. (2015). Judicial productivity, delay and efficiency: A directional distance function (DDF) approach. *European Journal of Operational Research, 240*(2), 592–601.

Falavigna, G., Ippoliti, R., & Ramello, G. B. (2018). DEA-based Malmquist productivity indexes for understanding courts reform. *Socio-Economic Planning Sciences, 62,* 31–43.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A (General), 120*(3), 253–290.

Ferrandino, J. (2014). Testing the Packer Theorem: The efficiency of Florida's criminal circuit courts. *American Journal of Criminal Justice, 39*(2), 375–393.

Ferrier, G. D., & Valdmanis, V. G. (2004). Do mergers improve hospital productivity? *Journal of the Operational Research Society, 55*(10), 1071–1080.

Finocchiaro Castro, M., & Guccio, C. (2014). Searching for the source of technical inefficiency in Italian judicial districts: An empirical investigation. *European Journal of Law and Economics, 38*(3), 369–391.

Finocchiaro Castro, M., & Guccio, C. (2016). Measuring potential efficiency gains from mergers of Italian first instance courts through nonparametric model. *Public Finance Review, 46*(1), 1–34.

Garden, K. A., & Ralston, D. E. (1999). The x-efficiency and allocative efficiency effects of credit union mergers. *Journal of International Financial Markets, Institutions and Money, 9*(3), 285–301.

Groff, J. E., Lien, D., & Su, J. (2007). Measuring efficiency gains from hospital mergers. *Research in Healthcare Financial Management, 11*(1), 77.

Grosskopf, S., & Valdmanis, V. (1987). Measuring hospital performance: A non-parametric approach. *Journal of Health Economics, 6*(2), 89–107.

Harris, J., Ozgen, H., & Ozcan, Y. (2000). Do mergers enhance the performance of hospital efficiency? *Journal of the Operational Research Society, 51*(7), 801–811.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis, 20*(1), 1–24.

King, G., Lucas, C., & Nielsen, R. A. (2017). The balance-sample size frontier in matching methods for causal inference. *American Journal of Political Science, 61*(2), 473–489.

Kittelsen, S. A., & Førsund, F. R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis, 3*(3), 277–306.

Kjekshus, L., & Hagen, T. (2007). Do hospital mergers increase hospital efficiency? Evidence from a National Health Service country. *Journal of Health Services Research & Policy, 12*(4), 230–235.

Kwoka, J., & Pollitt, M. (2010). Do mergers improve efficiency? Evidence from restructuring the US electric power sector. *International Journal of Industrial Organization, 28*(6), 645–656.

Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics, 84*(2), 205–220.

Lewin, A. Y., Morey, R. C., & Cook, T. J. (1982). Evaluating the administrative efficiency of courts. *Omega, 10*(4), 401–411.

Lynk, W. J. (1995). The creation of economic efficiencies in hospital mergers. *Journal of Health Economics, 14*(5), 507–530.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *National Institute of Science of India, 2*(1), 49–55.

Major, W. (2015). Data Envelopment Analysis as an instrument for measuring the efficiency of courts. *Operations Research and Decisions, 25*(4), 19–34.

Månsson, J. (1996). Technical efficiency and ownership: The case of booking centres in the Swedish taxi market. *Journal of Transport Economics and Policy*, *30*(1), 83–93.

Mattsson, P., Månsson, J., Andersson, C., & Bonander, F. (2018). A bootstrapped Malmquist index applied on Swedish district courts. *European Journal of Law and Economics*, *46*(1), 109–139.

Mattsson, P., & Tidanå, C. (2019). Potential efficiency effects of merging the Swedish district courts. *Socio-economic Planning Sciences*, *67*, 58–68.

Ministry of Finance. (2017). Budgetpropositionen för 2018 (*The budget bill 2018*), 2017/18:1, Appendix 4 the judicial system, Stockholm. (**in Swedish**).

Ministry of Justice. (2000). 1999/2000:106 *Reformeringen av domstolsväsendet—en handlingsplan (Reforming of the judiciary—an action plan).* Stockholm. (**in Swedish**).

Ministry of Justice. (2015). *The Swedish judicial system*. http://www.regeringen.se/49e9fa/contentassets/24ef36d9206a41fb80d6b74030fd9ef6/the-swedish-judicial-system.pdf. Accessed 20 April 2018.

Nissi, E., & Rapposelli, A. (2010). A data envelopment analysis of Italian courts efficiency. *Statistica Applicata Italian Journal of Applied Statistics, 22*(2), 199–210.

O'Donnell, C. J., Rao, D. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics, 34*(2), 231–255.

Odeck, J. (2008). The effect of mergers on efficiency and productivity of public transport services. *Transportation Research Part A: Policy and Practice, 42*(4), 696–708.

OECD. (2011). Impact evaluation of merger decisions. In *Organisation for Economic Co-Operation and Development*. DAF/COMP(2011)24. http://www.oecd.org/daf/competition/Impactevaluationofmergerdecisions2011.pdf Accessed 2 May 2018.

Ouellette, P., & Vierstraete, V. (2010). Malmquist indexes with quasi-fixed inputs: An application to school districts in Québec. *Annals of Operations Research, 173*(1), 57–76.

Papadimitriou, M., & Johnes, J. (2018). Does merging improve efficiency? A study of English universities. *Studies in Higher Education*. https://doi.org/10.1080/03075079.2018.1450851.

Pedraja-Chaparro, F., & Salinas-Jimenez, J. (1996). An assessment of the efficiency of Spanish Courts using DEA. *Applied Economics, 28*(11), 1391–1403.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33–38.

Rubin, D. B. (1980). Bias reduction using mahalanobis-metric matching. *Biometrics, 36*(2), 293–298.

Santos, S. P., & Amado, C. A. (2014). On the need for reform of the Portuguese judicial system—Does data envelopment analysis assessment support it? *Omega, 47*, 1–16.

Schmitt, M. (2017). Do hospital mergers reduce costs? *Journal of Health Economics, 52*, 74–94.

Schneider, M. R. (2005). Judicial career incentives and court performance: An empirical study of the German labour courts of appeal. *European Journal of Law and Economics, 20*(2), 127–144.

SFS (1996:242). *Lag om domstolsärenden (The Swedish Court Matters Act)*. Stockholm: Ministry of Justice. (**in Swedish**).

SFS (2011:203). *Budgetlagen, (The Swedish Budget Act)*. Stockholm: Ministry of Justice. *Chapter 1, 3§* (**in Swedish**).

Silva, M. C. A. (2018). Output-specific inputs in DEA: An application to courts of justice in Portugal. *Omega, 79*, 43–53.

Simar, L., & Wilson, P. W. (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis, 13*(1), 49–78.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics, 136*(1), 31–64.

Wångmar, E. (2003). Förändringar av Sveriges underrätter efter år 1900 (Changes in the lower courts in Sweden after 1900). *HumaNetten, 13*, 1–11. (**in Swedish**).

Worthington, A. C. (2001). Efficiency in pre-merger and post-merger non-bank financial institutions. *Managerial and Decision Economics, 22*(8), 439–452.

Wu, D. D., Zhou, Z., & Birge, J. R. (2011). Estimation of potential gains from mergers in multiple periods: a comparison of stochastic frontier analysis and Data Envelopment Analysis. *Annals of Operations Research, 186*(1), 357–381.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.